UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

Máster en Astrofísica

TRABAJO DE FIN DE MÁSTER

Construcción de un catálogo de regiones libres de emisión de CO en el plano Galáctico para mejorar la eficiencia de los observatorios submilimétricos.

Developing a catalog of CO emission-free regions in the Galactic Plane to improve efficiency of sub-mm observatories

Julián José Miranzo Pastor

Tutor:

Francisco Miguel Montenegro Montes

Curso académico 2023

Developing a catalog of CO emission-free regions in the Galactic Plane to improve efficiency of sub-mm observatories

Julián José Miranzo Pastor

June 7, 2023

Abstract

A recurrent problem faced by mm/sub-mm observatories in their daily operations is the identification of regions of the sky free from molecular emission, in particular when observing the densely populated regions around the Galactic plane. In the so-called total-power observations, the telescope needs to quickly switch between the science target and some close emission-free (OFF) position, so it is possible to remove the sky contribution from the faint astronomical source signal. This project aims at building up a catalogue of *CO* emission-free positions by analysing the temperature-calibrated data cubes of large-area surveys of the Galactic Plane (CfA, Planck) by developing an original code capable of discerning potential off-positions in the [+5,-5] degree Galactic latitude range.

Un problema recurrente al que se enfrentan los observatorios mm/sub-mm en su día a día es la localización de regiones del cielo libres de emisión molecular, en particular cuando se observan regiones más densamente pobladas cerca del Plano Galáctico. En las llamadas observaciones 'total-power', el telescopio necesita cambiar rápidamente entre la región de interés y una zona cercana sin emisión (OFF), de modo que sea posible eliminar la contribución del cielo de la débil señal del objeto astronómico. Este proyecto pretende construir un catálogo de regiones libres de emisión de CO analizando los cubos de datos calibrados en temperatura de mapas de gran área en el Plano Galáctico (CfA, Planck) mediante el desarrollo de código original capaz de discernir potenciales posiciones off en el rango de latitudes galácticas entre [+5,-5] grados.

Contents

1 Introduction

Millimetre and submillimetre astronomy began in the 1960's as the branch of astronomy conducted at mm/sub-mm wavelengths (between a few hundred microns and 3 mm). Its main objective is the observation of molecular clouds and dark cloud cores, thus trying to understand the processes behind star formation, determine chemical abundances, identify cooling mechanisms for different molecules or give information on the mechanisms for the formation and evolution of galaxies.

Molecular clouds are composed mainly of H_2 and atomic He. However, He transitions are far from excitable in the conditions of typical molecular clouds (10-20 K, 10^2-10^6 cm^{-3}) (Osterbrock & Ferland, 2006), and molecular hydrogen does not have a permanent dipole, so it cannot have rotational transitions. Fortunately, we can trace molecular clouds and interpret the Galactic emission at almost every other wavelength band with the carbon monoxide (CO) (Heyer & Dame, 2015), which is the second most abundant molecule in these clouds. The ^{12}CO transitions allow us to track down the density all over the cloud, except maybe in the denser regions, where the medium might be optically thick; here, one can use the ^{13}CO transitions, as the abundance of ^{13}CO is much lower, $[^{12}CO]/[^{13}CO] = 30 - 40$ (Blake et al., 1987), and the cloud is optically thin to these wavelengths. These conditions make CO and its isotopologues extremely valuable molecules to observe for astronomers.

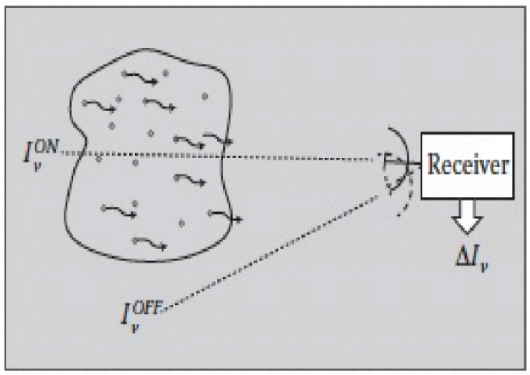
Observations from the ground in this spectral range are very challenging due to the need of very efficient antennas and sensitive receivers as well as a dry and stable atmosphere.

In this work, we focus on the problem with the sky emission. A peculiarity of sub-mm astronomy is that the atmosphere is much warmer (around 300 K) than the astronomical objects of interest (a few tens of K). This means that the emitted power by the atmosphere is several orders of magnitude larger than the signal received from the astronomical objects at these wavelengths. In addition, the excitation of certain molecules like H_2O, O_2, etc. in the atmosphere (Battistelli et al., 2012) produce rotational transitions that disturb these observations at certain wavelengths. This is the reason why mm/sub-mm observatories are built at very dry places (like Antarctica) and normally at high altitudes to get rid of a good fraction of the troposphere's emission and water contents. Because of this problem of a sky much brighter than the astronomical signal, typical observing modes are designed to try to subtract as efficiently as possible the sky emission.

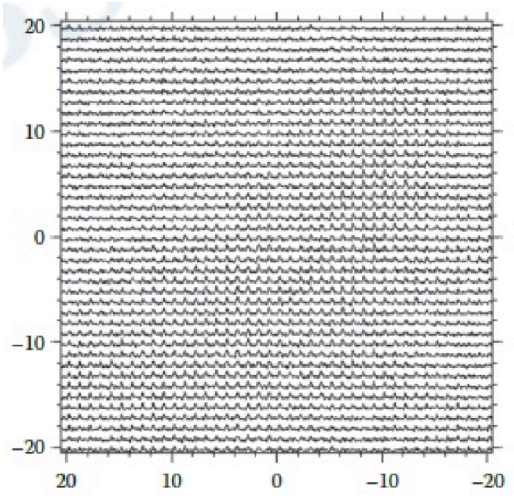
Absolute position switching (APS) is a standard observing strategy, specially useful for the observation of objects only a few beams across —this includes point sources and sources with limited extent—. In this observing mode, the telescope alternatively measures the intensity along the line of sight of two different positions; the first position, called ON-position, is the corresponding to the object of interest, while the second position, called OFF-position, is placed (ideally) at a close angular distance from the object, as shown in Figure 1.a. These intensities are actually flux densities, as they are measured with a telescope beam of some solid angle, and the detection system converts them into measurable powers. The powers of the ON and the OFF positions contain both the power related to the noise of the detection system and the emission of the atmosphere (in the case of ground telescopes and other kinds of telescopes not space-based). If we then subtract the two powers (ON-OFF), the remaining is the power associated with the astrophysical object (Walker, 2015). In order to achieve a successful suppression of the atmospheric emission, the ON and OFF positions should be close enough so that the atmospheric emission is the same in both measurements. In particular, the elevation of the positions should be as close as possible, as the variation in elevation means a different air mass, which leads to a different emission. The switch between the ON and the OFF position should be done frequently enough to avoid changes in the sky emission and system temperature between both integrations (typically less than 1 minute).

On-the-fly mapping is the preferred observing mode when carrying out studies involving the observation of extended sources (more than a few beams across) or when conducting large-area surveys. This technique requires the telescope being driven across a certain field of the sky while recording

data and the telescope's position continuously (see Figure 1.b), at a rate typically less than 0.1 seconds (Mangum, 2006). It presents several advantages over the APS method: telescope overhead is nearly eliminated, since no specific position needs to be acquired and we obtain useful data even when accelerating or decelerating the telescope; the decrease of the overhead leads to a rapid covering of the field, reducing changes in the characteristics of the atmosphere and changes in the system (antenna pointing, calibration, etc), and systematic changes can be averaged down and corrected with cross correlation techniques; there is a higher observing efficiency, reducing the RMS by a factor of approximately two with ≥ 100 ON measurements, and there is only one OFF measurement needed, with an optimal (theoretical) duration, $t_{off}^{optimal} = \sqrt{N} t_{on}$ (Mangum et al., 2007).



(a) Absolute position swithching (APS) mode. The telescope first measures the ON intensity and then changes the position to measure the OFF intensity right after.



(b) OTF map of $CO(3-2)$ emission in Orion made at the Arizona Radio Observatory. Spectra are taken as the telescope drifts across the object. Each row of spectra shares a common OFF-position.

Figure 1: Comparison between position-switching and OTF observing modes. Retrieved from (Walker, 2015).

For these observing modes, we need OFF-positions that do not contain the emission of another astronomical object, because the subtraction ON – OFF would result in a mixed signal of both objects otherwise. To avoid this contamination, we need to be sure that the OFF-position has no measurable emission associated, that is, the position is "clean". This problem is of particular importance in regions of the Galactic Plane where gas and dust from molecular clouds cover large areas of the sky.

2 Objectives

Millimetre and submillimetre observatories need to identify OFF-positions free from emission to be able to subtract the atmospheric emission in their routine observation modes (ON-OFF and OTF maps). The selection of these OFF-positions is usually done at the time of observation. It involves selecting some candidate OFF-position(s) close to the science target and making a short integration to verify the absence of emission. Having already a database of verified clean positions would certainly speed up this process, leaving more time to spend on the scientific observations. Sometimes, there are a few OFF-positions registered in the observatories —for example, the Atacama Large Millimeter Array (ALMA) has a catalog[I] with around 200 positions without CO emission along the Galactic Plane

[I] https://almascience.eso.org/sc/

and some other scientifically relevant regions—, but many sky regions do not have a suitable closeby OFF-position in this database, specially regions where observations are not very common. Moreover, the method of selection and categorization of the OFF-positions (in site) in big observatories is rarely systematized, making it a slow process every time a new OFF-position is needed.

The aim of this work is to develop an algorithm that is able to find new OFF-positions in the Galactic Plane (we will focus in latitudes between $b = \pm 5°$) that can provide many viable new candidates. Specially important will be to find regions far from the currently catalogued OFF-positions and close to the Galactic Center, an area of particular scientific interest and commonly observed. To do so, we have collaborated with the Joint ALMA Observatory (JAO) and are using the OFF-positions in their database as a reference, in order to have a reasonable criterion of selection.

The resulting catalog after the implementation of this algorithm is expected to be of high impact in mm and sub-mm observatories, as it can help in almost all routine observations, in particular near the Galactic Plane which concentrates a good fraction of the observing projects. In addition, the method should be easily extended to other potentially interesting regions.

3 Single-dish large-area *CO* surveys

For this work to be possible, the existence of public archival data of large-area surveys that cover the Galactic Plane has been crucial. It was also relevant that the surveys we used came from single-dish telescopes (with only one antenna) and not from interferometers like ALMA, IRAM/Plateau-de-Bure or the Submillimeter Array (SMA). The reason behind this is that diffuse emission is filtered out in the maps observed with interferometers, leaving "holes" at the bigger spatial scales. Only with single-dish observations this large-scale diffuse emission can be recovered. This is known as the short-spacing or zero-spacing problem (Stanimirović, 2002).

During the development of this work, mainly 3 different single-dish large-area surveys have been explored to determine potential positions without emission in the CO transition: SEDIGISM, the CfA CO survey and the Planck all-sky survey.

3.1 SEDIGISM survey

The SEDIGISM (Structure, Excitation and Dynamics of the Inner Galactic Interstellar Medium) survey (Schuller et al., 2021) mapped 84 deg^2 of the Galactic Plane, between $l = 300°$ and $l = 31°$, using the APEX (single-dish) telescope. Several molecular transitions were observed between frequencies 217-221 GHz, the prime target lines being $^{13}CO(2-1)$ and $C^{18}O(2-1)$.

The observations were done using on-the-fly mapping mode, with an angular resolution of $30''$ and a typical noise level of order 1 K at 0.25 km s^{-1}. In these observations, the reference OFF-positions were situated at $\pm 1.5°$ off the Galactic Plane in most cases. The gross of the survey covers the Galactic Plane in a range of latitudes of $b = \pm 0.5°$, though it could be extended to $b = \pm 1°$ for the Galactic Center ($358° \leq l \leq 1.5°$).

The data is arranged in 78 one-degree-longitude cubes, centered every 0.5° in longitude, meaning each cube overlaps a combined total of 1 degree of its two adjacent cubes. The data-cubes are calibrated in temperature, centred on the most relevant spectral lines possibly detected in the band-pass, and projected on $9.5''$ pixels, and for the ^{13}CO and $C^{18}O$, the velocity resolution is 0.25 km s^{-1}, and 0.5 km s^{-1} for the rest of the transitions. The RMS is slightly different over the whole survey, with values between 0.5 K and 1.6 K, but with most of the fields below 1.0 K.

Based on the SEDIGISM data cubes, Duarte-Cabral et al. (2021) created a catalog of dense molecular clouds in our galaxy, which was done following a technique of hierarchical clustering to find clusters of ^{13}CO emission both in spatial coordinates and in velocity. For this purpose they developed the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm, which uses the `astrodendro` library from Python.

3.2 CfA *CO* survey

The Center for Astrophysics (CfA) *CO* survey (Dame et al., 2001), is a large-scale composite *CO* survey that covers the entire Galaxy by combining the *CO* surveys obtained with the CfA 1.2m telescope together with other 31 surveys obtained with that same telescope and the very similar one on Cerro Tololo (Chile) as shown in Figure 2. The surveys were integrated using moment masking, in order to produce composite complete maps of the Milky Way with little noise, providing detailed information about the main structure of the Galaxy and individual clouds.

The spectra in positions close to the Galactic Plane were obtained by position switching every 15 seconds between the ON-position and two OFF-positions, depending on the elevation of the ON position. There were residual offsets of less than 1 K but were removed by line fitting the emission-free ends of the spectrum.

The beams of the antennas are very similar, with Half-Power Beam-Width (HPBW) values close to 8.5'. The RMS noise and velocity resolution remain almost constant for each of the surveys, varying between 0.12 K and 0.43 K, as the integration time of each scan was adjusted depending on the system temperature determined from 1 second calibrations. The integration times also depend on the target RMS, and vary from 20 seconds to 5 minutes.

The data was stored in two 3D (velocity, longitude, latitude) `.fits` cubes, each of one covering a half of the Galaxy. Also, several versions of the data set and of its subsections (individual surveys) were publicly distributed.

3.3 Planck 857 GHz survey

The Planck satellite from the European Space Agency scanned the sky in the microwave and submillimetre ranges, producing nine all-sky maps in frequency bands from 30 to 857 GHz, centered in 30, 44, 70, 100, 143, 217, 353, 545 and 857 GHz (Planck Collaboration et al., 2020a). The Planck 857 GHz map in particular was carried out with an effective beam of 4.22' (HPBW) and an average noise level of 0.72 kJy/sr.

The Planck maps include emission from several sources: the Cosmic Microwave Background (CMB), low frequency/synchrotron, *CO* lines, compact objects and thermal dust, and this combined emission needs to be differentiated in order to make a usable *CO* map (Planck Collaboration et al., 2020b). The total emission measured at a certain frequency is modeled as a weighted sum of all of the emitting sources, where the CMB and the *CO* lines emissions add linearly to the total value, and the low frequency/synchrotron, the compact objects and the thermal dust emission are weighted with a power of the frequency (see Planck Collaboration et al. (2014) and Annex A.2 from Planck Collaboration et al. (2020b) for a more detailed explanation of this model).

The primary goal of the Planck mission was to obtain deep high-resolution images of the CMB, which would and have confirmed the success of the ΛCDM cosmology model, but also have determined with better precision five of the six parameters of this model (Planck Collaboration et al., 2020a). Nevertheless, our goal with this work stays far from this, and will only make use of the *CO* derived maps extracted from the data from Planck observations.

4 Selection of potential positions

In this section we will explain the process of developing a method for selecting suitable positions for our catalog. We follow a chronological order, since the first strategy did not go as planned and made us rethink all the method from the start, including new crucial restrictions which we did not consider the first time.

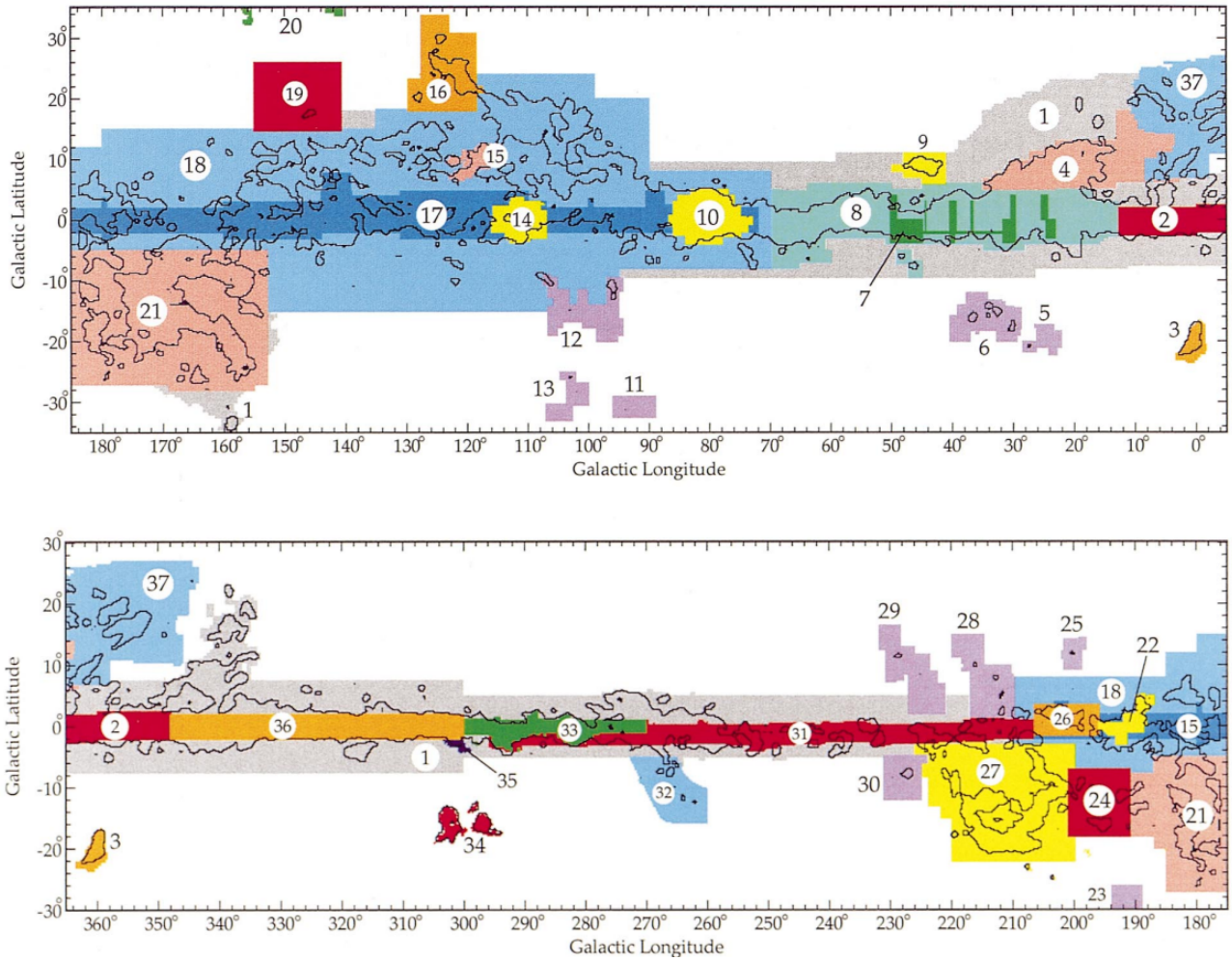


Figure 2: Areas covered by the individual CO surveys used to synthesize the composite survey of the Milky Way, numbered as shown in Table 1 from (Dame et al., 2001). A single contour at 2 K km s^{-1} from a smoothed version of the velocity-integrated map is overlaid. This figure was retrieved from (Dame et al., 2001).

4.1 First selection, using the SEDIGISM survey

Our first selection was based on the SEDIGISM datacubes. We explored various strategies on this dataset and finally we selected a few candidates. As explained below, we could have some of these positions observed by ALMA, for verification and to test if the method was valid for OFF-position selection, finding a negative result.

4.1.1 Exploring different strategies

Our first and most ambitious approach to developing this catalog came with the idea of using the `astrodendro`[II] package, following the work in Duarte-Cabral et al. (2021) with the datacubes of SEDIGISM. Our strategy was to invert these dendrograms so we could identify those regions with the lower emission —ideally without emission—, contrary to what was done in Duarte-Cabral et al. (2021), which was to find the regions with significant emission. The advantage of using the full 3D data cubes is that we could in principle propose candidate positions usable in different velocity ranges than those of the CO emission detected in the survey. This could be particularly interesting in crowded regions close to the Galactic Center, something that might be interesting for some works. A drawback of this

[II] https://dendrograms.readthedocs.io/en/stable/

method is the long time the algorithm needs to compute the dendrograms together with the resources it demands computationally speaking.

A second idea was to integrate in velocity the SEDIGISM data cubes and build up 2D integrated images where we could look for OFF-positions free from emission at all velocities. The benefit of using this method is that the candidates should work in the whole range of velocities covered by the survey (± 200 km s^{-1}), meaning that observations of objects with different radial velocities in the same region of the sky could make use of the same OFF-positions, regardless of their velocity. The disadvantage, on the other hand, would be the lose of all the positions with localised emission in a small range of velocities which could be used with the previous method. However, the technical difficulties we were having with the dendrograms and the limited time we had encouraged us to choose this second method.

Another possibility could have been to go over all spaxels in the 3D cubes (as we did with the 2D moment-integrated cubes) and find those pixels without emission, in order to find the velocities in which there is no ^{13}CO. But the time needed to execute this would have been extremely prolonged, which made us not even ponder this option.

4.1.2 Selection method with SEDIGISM

To begin with, we had to integrate the cubes in velocity to be able to analyse the data in a 2D image. To do this, we made use of the tool CARTA[III], a very easy-to-use interface with which we could do the integrations straightforwardly.

Once the cubes were integrated, the easiest —not necessarily the best— way to predict if a certain position is a good candidate to be an OFF-position is to check the emission in the pixel corresponding to this position: if the emission in the pixel is within the noise of the image, the position should not be contaminated (at least not with an emission higher than the noise level of the cube/region).

We need to consider that the 2D images have a certain RMS noise and we cannot know if there is some residual emission lower than this value. There is not much we can do about this, but we need to have a criterion of selection for the positions with emission within the noise.

A possible criterion (the one we used) is to select the local minima under a certain value, which in our case was an intensity of 0 K. But another important factor to consider is the extension of the OFF-position that we want to compile. The SEDIGISM data maps have a pixel size of $\sim 1/3$ of the APEX telescope beam, according to Nyquist sampling. For the OFF-positions to be useful for dishes with different diameters and at different resolutions, we need that all the area covered by the antenna beam is clear from CO emission; therefore, we will need to check the emission in all pixels within one beam. ALMA beam is $19''$ at 300 GHz for a 12-m antenna and $33''$ for a 7-m antenna (ALMA, 2023), while SEDIGISM cubes have a pixel size (in the spacial directions) of $9.5''$, then a region of 5px $\times$ 5px without emission would be a convenient size with which to work, having in mind that also smaller antennas might also want to make use of the OFF-positions.

The algorithm to search the positions is quite simple, but considers all of these issues and tries to solve them as explained before. In short, the algorithm goes through all of the pixels of the 2D map and checks if the emission of each pixel is lower than zero. If it is, then it looks if all of the 24 other pixels in a 5px $\times$ 5px ($47.5'' \times 47.5''$) box around it also have emission lower than zero. When this is also true, then the position is stored as a potential OFF-position in a list.

To try out if the method worked as expected, we run the algorithm over the regions 000 and 307, corresponding to the moment-integrated cubes of the regions $l = \pm 1°, b = \pm 1°$ (centered in $l = 0°$) and $l = 306° - 308°$, $b = \pm 0.5°$ (centered in $l = 307°$). The results, shown in Figure 3, seemed promising: positions tend to appear more towards higher values of b, specially in the region 000, or concentrated in small clusters located in low emission regions. There are also no positions in the areas with more

[III] https://cartavis.org/

gas emission, as we can see in the right side of Figure 3.b or in the central horizontal section of Figure 3.a, where the emission of the clouds is significantly higher than in the rest of the cube.

We also developed a less restrictive method, in case we could find more candidates even closer to the Galactic Center, something that would be very useful if they turned out to be valid OFF-positions. This second algorithm only checked 4 additional positions once it found a non-emitting pixel, each of them 5px to the right, left, top and bottom from the initial one. The results from this method were quite similar but with a lot more of candidates, as expected from the lower constraints demanded.

Once the algorithm(s) seemed to work as expected and produced the first list of candidate regions, and before extending the work to all the cubes, we sent some of the positions to ALMA, to check if the method(s) of selection could find valid OFF-positions near the Galactic Plane.

4.1.3 Observations with ALMA and results

In December 2022, we presented the project to the JAO/ALMA team in one of their meetings while we were still developing the selection method. As they found our work useful for their observations, they offered us some observing time while ALMA was in their shutdown period[IV]. From the catalog of possible OFF-positions in the 000 and 307 cubes, we handpicked a total of 28 candidates, eighteen from the 000 region and the other ten from the 307 region. The selection was made visually, in order to include positions of all kinds (closer and farther from highly emitting areas in the cubes and also from the galactic plane).

We also set priorities to these positions, shown in Figure 4. We define as high priority candidates (they appear in white in Figure 4) the positions close to dense regions found with the strong constraints method, explained in Section §4.1.2. These candidates are not only the most interesting ones due to their closeness to the Galactic Center and high(er) emission regions, but also because, as a consequence of this, they are also the most likely to be contaminated; if they do not show contamination it is highly probable that candidates in less dense regions do not show contamination either.

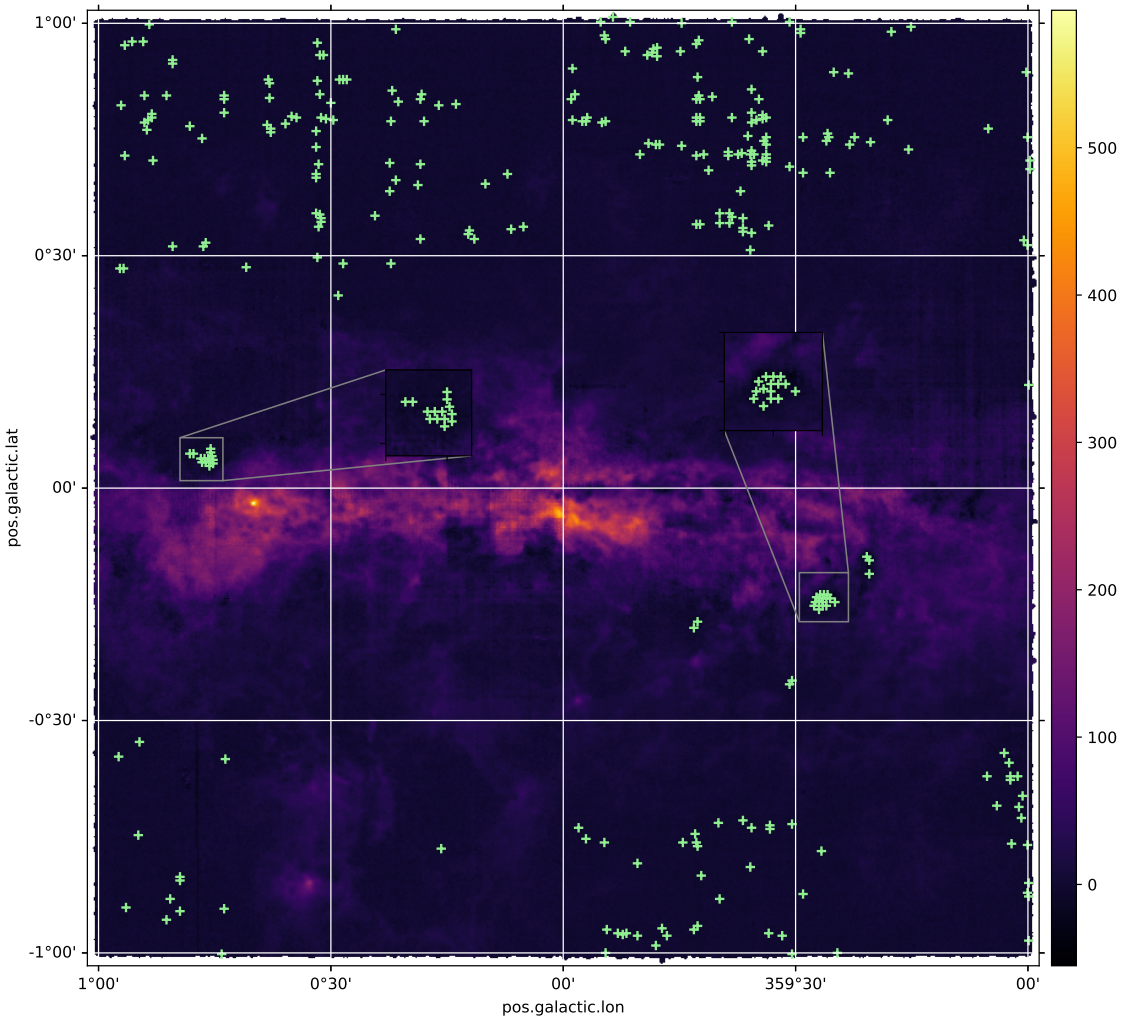
The mid-high priority candidates (which are shown in blue in the figure) are similar to the first ones but this time they have been obtained with the low constraints method (see Section §4.1.2). The problem with this type of positions is that the closeness to dense regions together with a not so reliable method of selection increases highly the risk of them being contaminated. However, in the improbable case of them being viable OFF-positions, the chances of finding valid OFF-positions would increase considerably, as the number of positions determined with the low constraints method is significantly higher than the number of positions result of the strong constraints method.

Positions with mid-low priority (in yellow in the figure) are candidates obtained with the strong constraints method that we can find in farther and less dense regions of the cubes. Although a positive outcome from the high priority candidates could be enough to test if the method works, we need test it as much as possible to be sure if the candidates are valid in regions where the difference between emission and no-emission is lower.

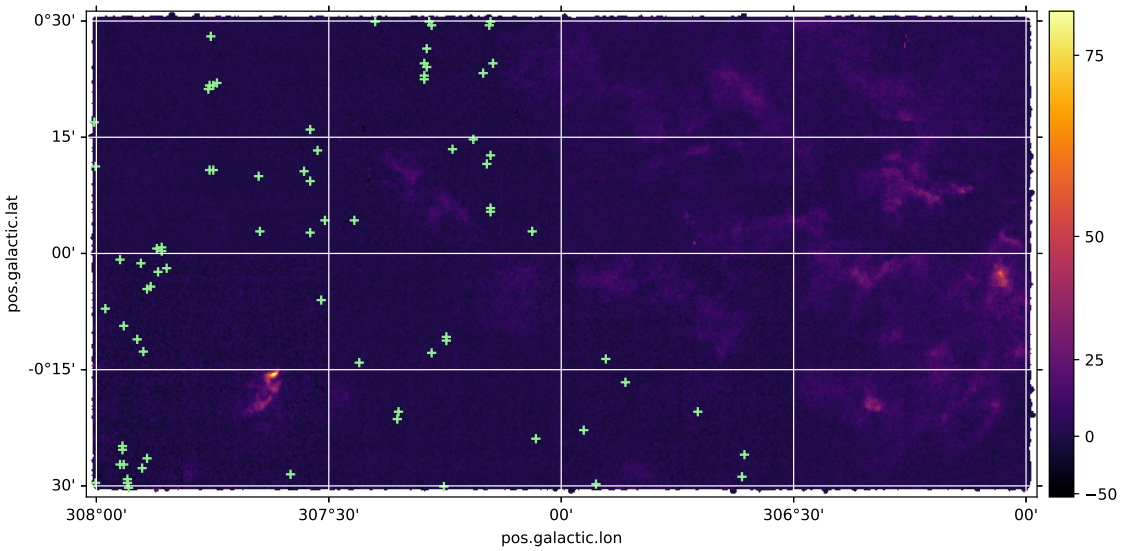
Finally, the low priority candidates (which are shown in purple) are positions very close to other high priority positions but located in the edges of the larger less dense regions where they are placed, in which we can find several clustered OFF-positions (see zoomed regions in Figure 3.a). Testing this candidates allows us to know if the method is valid only in the inside of large clusters of possible OFF-positions or if the candidates in the edges are potentially valid too.

Between the 3rd and the 4th of March, 2023, ALMA observed a first group of potential OFF-positions, corresponding to those in the 000 region. The results were discouraging, as all the candidates but one showed contamination in $^{13}CO(2 \rightarrow 1)$ (a table with the emission in each of the positions is available in the Appendix A).

[IV] The shutdown period is annually repeated in the same dates due to the worsening of the climatic conditions in the Llano de Chajnantor, where the observatory is located, during the summer period, a phenomenon known as "altiplanic winter".

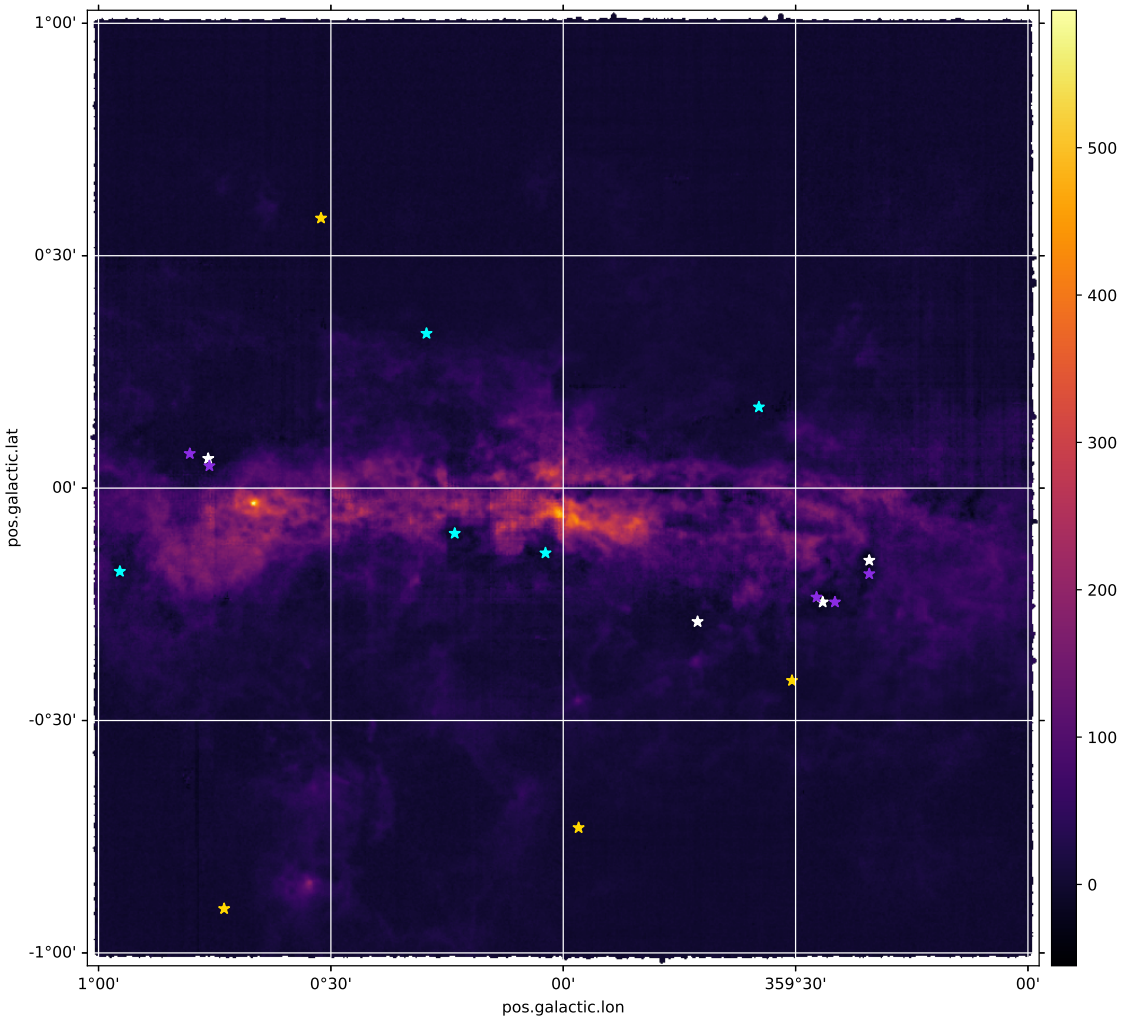


(a) Region 000

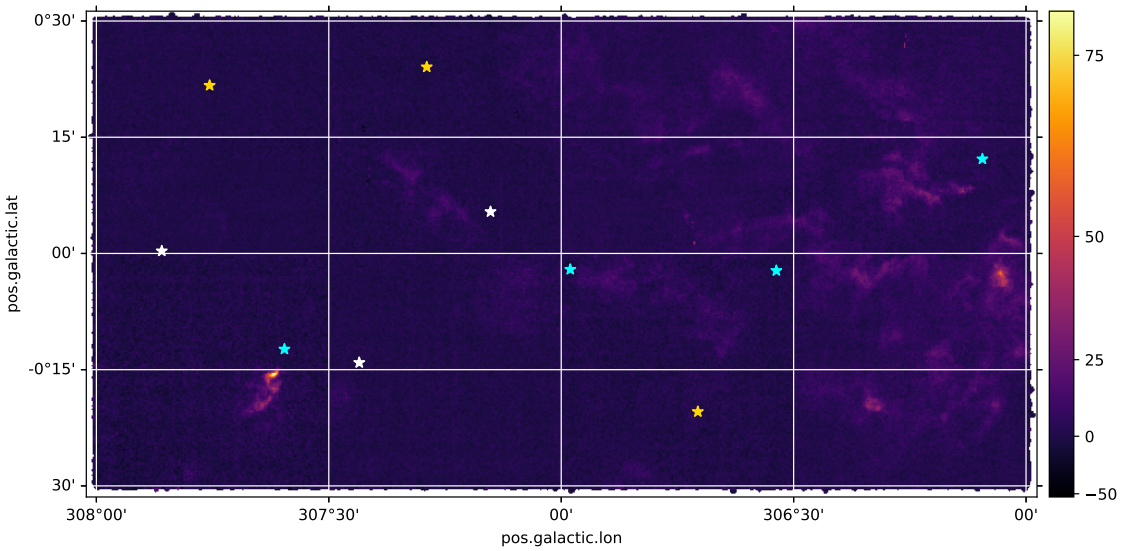


(b) Region 307

Figure 3: List of potential OFF-positions obtained with the low constraints method of selection. The map shows the emission of the Galactic Plane in K. The green crosses represent the candidates. In region 000, the majority of the candidates appear farther from the Galactic Plane ($|l| \gtrsim 30'$) or concentrated in small clusters located in no emitting areas of the cube, as shown in the zoomed squares. In region 307, we found remarkably less candidates than in the region 000, partly because of the restriction in latitudes (now only $l = \pm 30'$), and they tend to appear towards the left side of the region because of same gas concentration in the right side, as we can see by the color code.



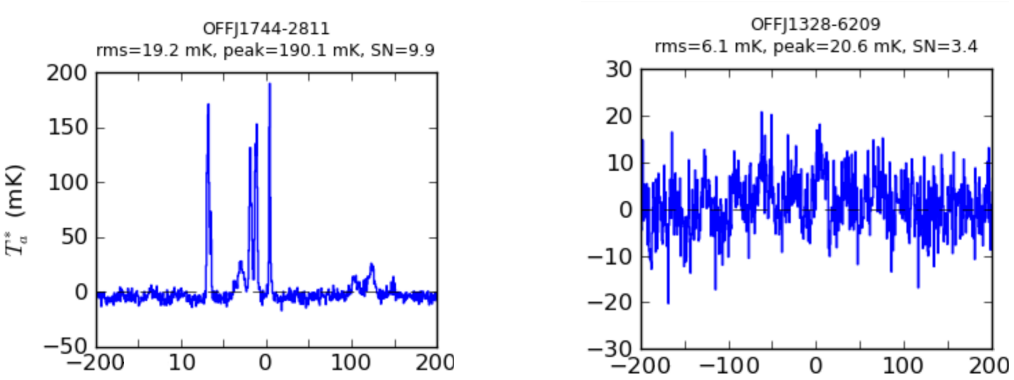
(a) Region 000



(b) Region 307

Figure 4: Candidates sent to ALMA for observation. The map shows the emission of the Galactic Plane in K. Color code: White - High priority. Strong constraints, unique regions, close to the denser regions of each cube or to the galactic plane. Blue - Mid-high priority. Low constraints, unique regions, close to the denser regions of each cube but with no near off-positions obtained with strong constraints. Yellow - Mid-low priority. Strong constraints, unique regions, close to less dense regions. Purple - Low priority. Strong constraints, repeated regions, edges of larger empty regions with several clustered off-positions (only for G000).

Before observing the OFF-position candidates with the ALMA Total Power (TP) array, the observatory staff, as a normal procedure for these observations, pre-checked the candidates with the $^{12}CO(J = 1 \rightarrow 0)$ data taken by the CfA *CO* survey (Dame et al., 2001), and *CO* and dust maps by Planck. All potential OFF-positions revealed some emission in the $^{12}CO(1 - 0)$ data, so it did not make much sense to re-observe these targets at the same frequency (115.271 GHz), as they were already showing emission in the survey. Still, ALMA observed the candidates in $^{13}CO(2 \rightarrow 1)$ (220.399 GHz), the original frequency at which these positions were selected from SEDIGISM. Unfortunately they all (except one) showed contamination in this transition too. The emission measured with ALMA toward one of the contaminated candidates and in the clear one is shown in Figure 5.



(a) CO contaminated candidate

(b) CO emission-free candidate

Figure 5: Emission measured by ALMA in given OFF-positions from our set of candidates chosen with the first selection method (using SEDIGISM). The X-axis represents the velocity range in km s^{-}1; the Y-axis shows the measured antenna temperature. These images were provided by ALMA.

4.2 Second selection, using CfA and Planck surveys

After the bad news of having all positions (except one) contaminated, we re-thought the method we were using to find them. Because all our candidates were within the noise of the SEDIGISM survey, we concluded that the survey was not deep enough in order to find empty regions, and that we had to choose a new survey (or more) in which to base the study.

4.2.1 ALMA off-position boundaries

As mentioned before in Section §4.1.3, to estimate the ^{12}CO emission in our candidates prior to observations, ALMA used the Dame et al. (2001) and Planck Collaboration et al. (2020b) surveys. They checked the emission in the coordinates of our candidates and, if the maps showed contamination, they did not observe the ^{12}CO transition —they did observe the ^{13}CO transition though—. Also, we learnt ALMA uses this same procedure when looking for new OFF-positions themselves. These facts made us consider that the CfA and the Planck surveys could be very useful to develop a new selection method of OFF-positions.

Using the CfA *CO* survey to find new candidates has some advantages and disadvantages. The most notable thing about using this survey is that, as we just explained, it is a ^{12}CO survey, which is more abundant and has more probable transitions than ^{13}CO. This makes ^{12}CO a better tracer of *CO* than ^{13}CO, which is a better tracer of *CO* only in the denser regions.

The dimensions of the pixels are bigger than in SEDIGISM, as it is also the area covered by the survey. This allows us to extend our catalog from the initial ±0.5 degree latitude range to the ±5 degree latitude range. This range could have been extended even more, but then the work would have started to move away from its initial purpose of finding OFF-positions near the Galactic Plane.

Something else to consider is that, as we explained in Section §3.2, the map is constructed as a mix of several surveys with different noise levels. However, we will see that this does not translate in a significant issue when we evaluate data from this map.

We then decided to use the Planck *CO*-maps (derived products) as our second criterion to determine these potential OFF-positions, from which we retrieved 36 sections of 10 degree × 10 degree along the $b = 0°$ axis, covering a total area of 360 degree × 10 degree ($l = 0° - 360°$, $b = \pm 5°$).

To work on a more reliable selection method, we decided to make use of the database of ALMA OFF-positions, which are *bone-fide* clean positions. We obtained from the CfA data maps the fluxes of the OFF-positions in ALMA's database and compared them with the SEDIGISM-based candidates' emission in the same CfA maps. In Figure 6, where we plot CfA vs Planck intensities we can see there is a clear distinction between the two sets of positions. The ALMA OFF-positions show smaller CO emission by several orders of magnitude, except for the fainter SEDIGISM-based candidates. Having this comparison in mind we will develop the new selection method in Section §4.2.2.

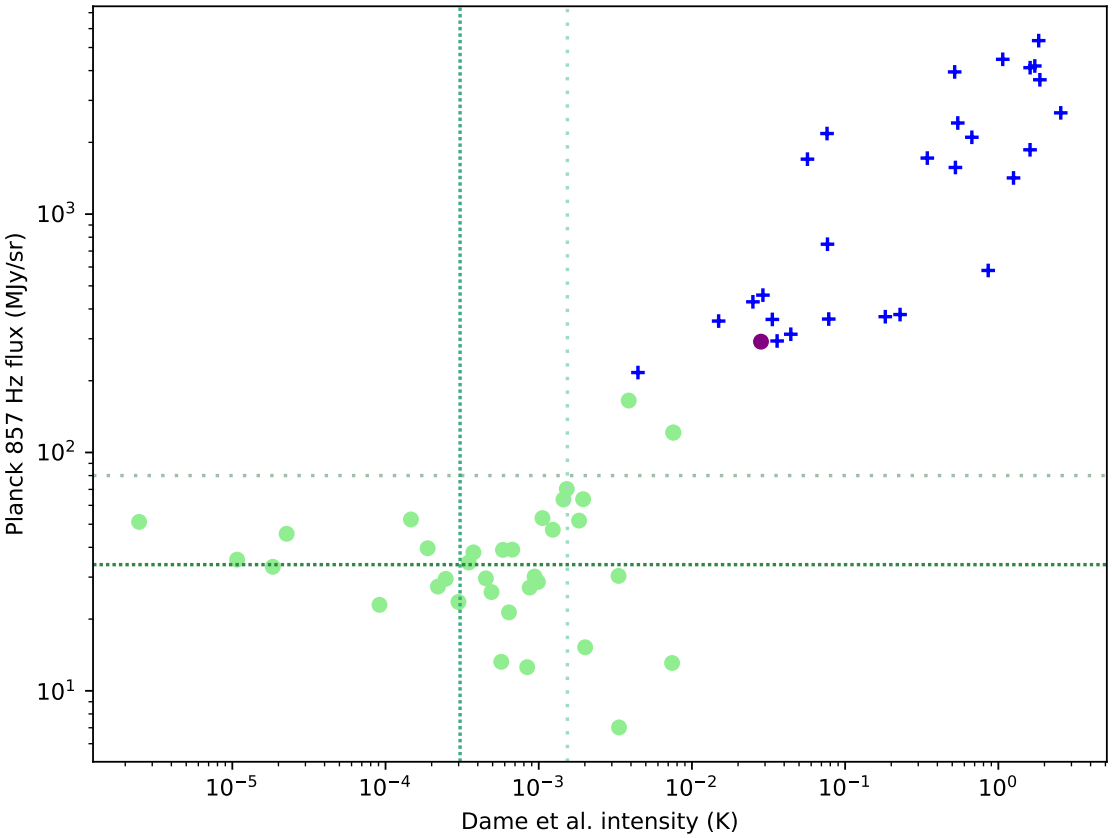


Figure 6: Emission in the Planck 857 GHz map vs emission in the CfA *CO* map. We have represented ALMA's database positions (green dots) together with our first 28 candidates (blue crosses), including the candidate which showed no emission in the ^{13}CO line (purple dot). The dotted lines represent the parameters of the distribution of emissions of the ALMA OFF-positions respective to each of the surveys (the horizontal lines correspond to Planck and the vertical lines to the CfA survey). The densely dotted lines represent the median of the distributions, while the loosely dotted ones represent 1 sigma deviation for CfA *CO* survey and a flux of 80 MJy/sr for Planck 857 GHz survey.

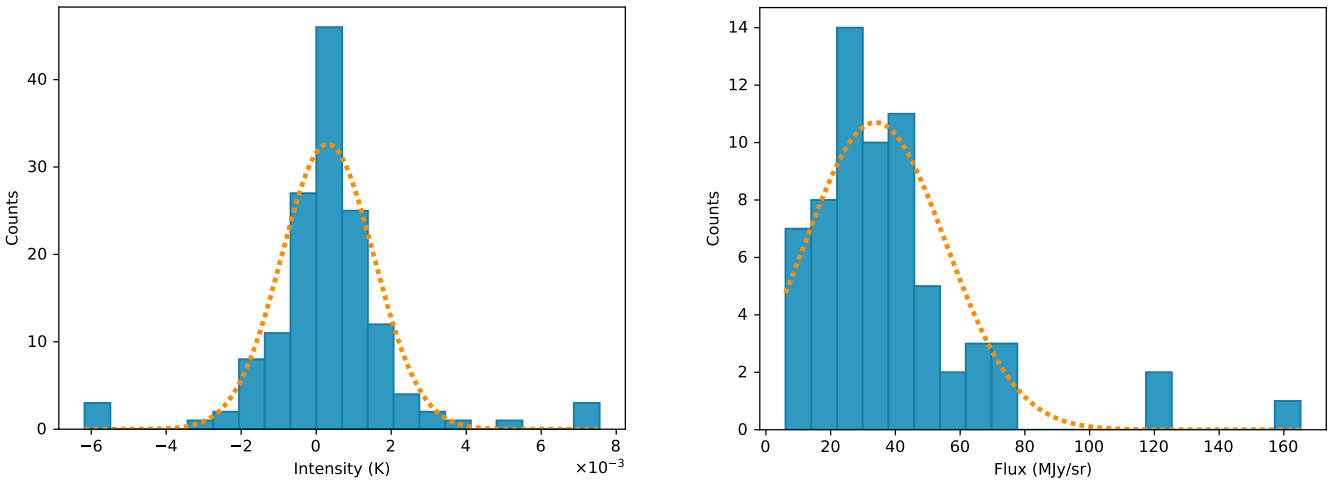
Having discarded our first candidates, we focused our attention to the ALMA OFF-positions, which are supposed to be free from CO emission. We looked at the distribution of pixel values in the CfA and Planck surveys to have them as reference for our new selection method. Looking at the histograms in Figure 7 we can see that the distribution could be compatible with a certain normal distribution except maybe for a few outliers (intensities farther than ±0.004K),

To define selection criteria based on these distributions, we will parametrize them using the median of the intensities instead of the mean value and the inter-quartile range instead of the standard deviation. By doing this, the parameters of the distribution of the intensities in CfA are:

$$\mu_{Dame} = 3.066 \times 10^{-4} \text{ K} \; ; \; \sigma_{Dame} = 1.232 \times 10^{-3} \text{ K}$$

On the other hand, the fluxes of the ALMA OFF-positions in the Planck 857 GHz survey do not seem to follow a normal distribution very well. Although we do not have many points in the range of latitudes $b = \pm 5^{\odot}$ (a total of 66 candidates), they should be enough to follow a normal distribution. We will parametrize this distribution in a similar way as we parametrized the data extracted from CfA (median, interquartile range) as we see in Figure 7.b, in this case not only to get rid of the outermost values and have a better estimator than the standard deviation, but also to have a better statistic of the distribution than the mean —which could have worked with the previous data, but does not work with the fluxes distribution from Planck—. The parametrization of the fluxes is then:

$$\mu_{Planck} = 33.84 \text{ MJy/sr} \; ; \; \sigma_{Planck} = 21.81 \text{ MJy/sr}$$



(a) CfA *CO*

(b) Planck 857 GHz

Figure 7: Emissions of the ALMA OFF-positions in CfA *CO* and Planck 857 GHz surveys. The histograms show (a) the distributions of intensities in CfA and (b) the distribution of fluxes in Planck 857 GHz of the OFF-positions of the ALMA catalog. The dotted line represents a normal distribution with the median and the interquartile range as parameters (instead of the mean and the standard deviation).

In Figure 7, we have plotted a normal PDF centered at the median values and with FWHMs equal to the distributions interquartile ranges. We can see that the CfA data is more consistent with such normal distribution.

4.2.2 Final selection method with CfA+Planck

As with the previous SEDIGISM-based selection method, we had to integrate in velocity the cubes of the CfA *CO* and the Planck *CO* surveys. For the CfA cubes (Harvard CfA, 2023) we used CARTA to integrate them in velocity, retrieving two 2D `.fits` files: one covering $l = 0^{\circ}$ to $l = 180^{\circ}$ and the second from $l = 180^{\circ}$ to $l = 360^{\circ}$, in a wide range of latitudes, which we will restrain to $b = \pm 5^{\circ}$.

For the Planck CO cubes, we used the online tool SkyView[V] to extract directly the 2D data from Planck. With it, we obtained the 36 cubes we mentioned in Section §4.2.1, with dimensions $10^{\circ} \times 10^{\circ}$,

[V] https://skyview.gsfc.nasa.gov/current/cgi/titlepage.pl

along the horizontal axis of the Galactic Plane (covering it completely). The cubes were retrieved with a pixel resolution of 5 arcmin / pixel and were tilted so the horizontal axis of the image matched the horizontal axis of the Galactic Plane.

After obtaining the integrated maps we need to keep in mind that the pixels are significantly larger than in SEDIGISM (it is done with a much smaller antenna) so it is not necessary to look for larger areas of free-emission.

The new algorithm we developed starts looking in both the CfA and Planck survey datasets simultaneously. The grid of explored positions is based on the center of each of the larger pixels in the CfA maps. Because the sizes and spacing of the pixels in both surveys differ, to calculate the corresponding flux in the Planck survey we considered the average flux in the 4 neighbour pixels (to give an example, if a certain position falls in the edge between two pixels, it would not be very rigorous to only consider the emission of one of them).

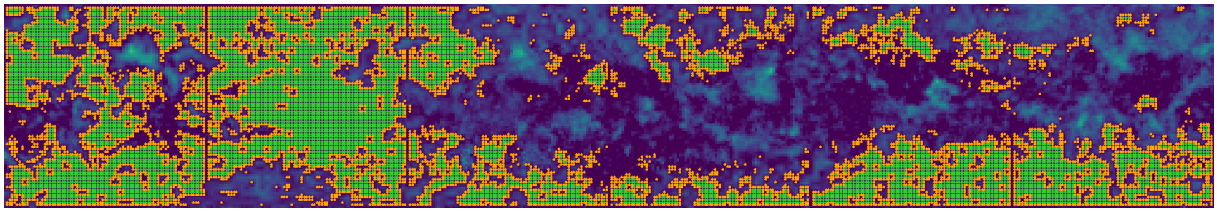
Now is where all the analysis we did in Section §4.2.1 comes into play. For the CfA *CO* survey, we wanted to be sure that there was no emission from ^{12}CO, so we only considered positions with intensities lower than $I = \mu_{Dame} + \sigma_{Dame}$ as possible candidates. This threshold keeps us far enough from the fainter positions in our first selection —see the leftmost blue cross in Figure 6, which is still far from the dotted vertical line that represents 1 σ in the CfA survey—. For the Planck CO map, as the data does not really follow a normal distribution, we decided to consider fluxes below the median of the distribution, $F = \mu_{Planck}$, even if this leaves several potentially good candidates out of the catalog. We can also consider all fluxes below $F = 80$MJy/sr, as this is the highest value we can find in the gross of the distribution from ALMA OFF-positions in Planck (Figure 7.b), though this is a riskier approach.

With these considerations, we run the algorithm and obtained two sets of potential OFF-positions: one with the more restrictive constraints and another with the less restrictive ones, but only in the region from $l = 300°$ to $l = 50°$ through zero. We then categorized these positions attending to the risk of them being or not valid OFF-positions (low/high risk): a low risk candidate is a pixel —a position— only surrounded by pixels that also are potential OFF-positions. In addition, we also classified the positions attending to the distance to an already existing OFF-position in the ALMA catalog, a property that we have called priority (high/low priority). Any candidate beyond 1 degree from an existing ALMA OFF-position is considered as high priority.

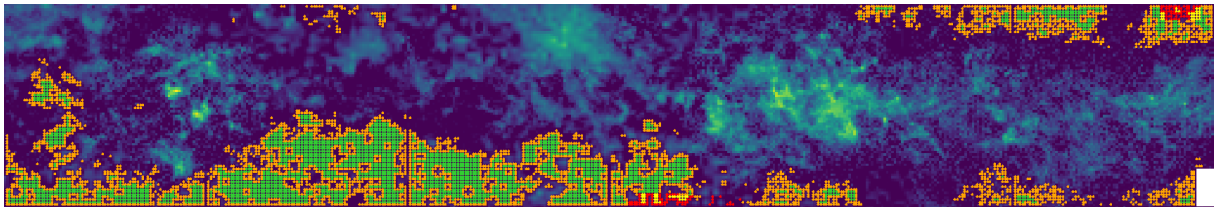
These categorized lists of candidates have been sent to JAO/ALMA for verification at the time of writing this work. However, given the busy period of observing programs in the winter season, these will probably take time to be checked. We are confident that these new positions, compatible with the noise distribution of the bona-fide OFF-positions in the ALMA database will yield the expected results.

5 Conclusions

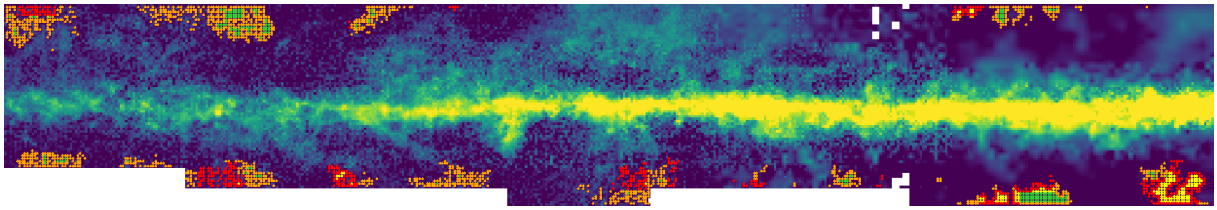
In this work we have compiled a catalog with potential OFF-positions near the Galactic Plane for the *CO* transitions. We hope the goodness of the method to be soon confirmed with ALMA technical observations. The restrictions imposed for the selection of the candidates are compatible with the existing ALMA clean OFF-positions. Moreover, they do not show contamination in the CO surveys we have used to determine their emission levels. Therefore we believe that the positions proposed have high chances of being valid OFF-positions for ALMA and other mm/sub-mm observatories. This belief is uplifted by several ALMA OFF-positions showing up in regions were a large number of candidates are found with our selection method, in many different areas of the Galactic Plane.



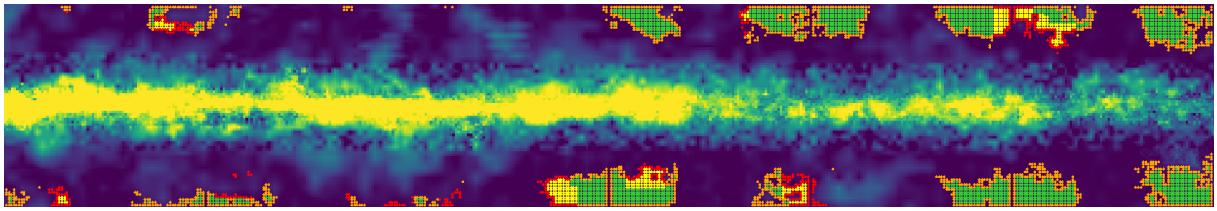
(a) Longitudes: 180° — 120°



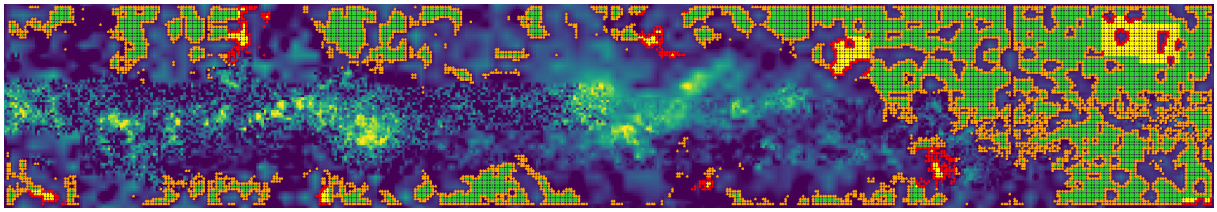
(b) Longitudes: 120° — 60°



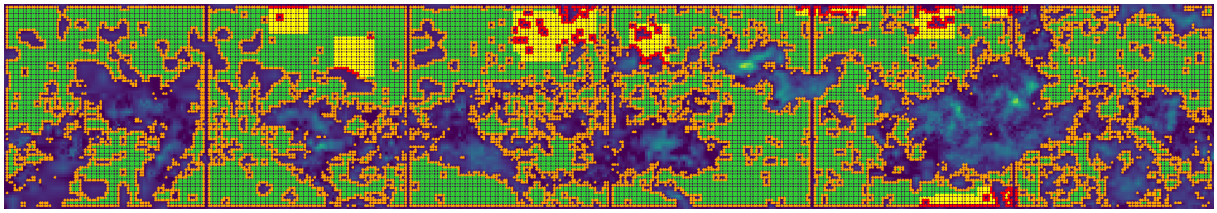
(c) Longitudes: 60° — 0°



(d) Longitudes: 360° — 300°



(e) Longitudes: 300° — 240°



(f) Longitudes: 240° — 180°

Figure 8: Visualization of the catalog of potential OFF-positions developed with the more conservative method of selection, from CfA *CO* survey and Planck CO data, that extends over all the Galactic Plane. The background represents the integrated emission of the Galactic Plane from the CfA maps. The map levelling is in logarithmic scale to enhance the contrast (yellow being the more intense and blue the less intense). The Y-axis goes from $b = -5°$ to $b = +5°$ in all the sub-sections. The positions of the catalog are represented as colored points. The color code of this positions is the following: GREEN - High priority, low risk; YELLOW - Low priority, low risk; ORANGE - High priority, high risk; RED - Low priority, high risk. Risk and priority are as defined in the text.

Pending the verification by ALMA of some of our candidates, we expect this catalog to become a very useful tool for mm and sub-mm observatories at the time of finding suitable OFF-positions, being able to reduce considerably the time spent looking for them. Our intention is to publish the list of OFF-positions and then develop a user-friendly tool to facilitate the search of positions near a specific astronomical object.

The code with the algorithm used to determine these candidates will also be publicly available. Modifications can be easily done in order to extend the areas covered by the algorithm, specially in scientifically interesting regions. Moreover, although the code does not work directly with other surveys, it should not be complicated to adapt it with a few changes, what could extend the catalog for other transitions.

Acknowledgments

I wanted to thank Dario Colombo and Ana Duarte-Cabral for all their help understanding the functioning of `astrodendro` and the data they reduced (and offered us) for the dendrograms, which in the end we did not use to produce our final catalog. I am also very grateful for the support from the JAO team, specially to Celia Verdugo, and for the opportunity we were given to have our candidates being observed by ALMA in February and the generous offer of new candidates being observed again in a near future. Finally, a very special thanks to Francisco Montenegro for directing this work and for being of such valuable help during all the process of developing this project.

References

ALMA. 2023, About Atacama Large Millimeter Array, https://almascience.eso.org/about-alma/alma-basics

Battistelli, E. S., Amico, G., Baù, A., et al. 2012, Monthly Notices of the Royal Astronomical Society, 423, 1293, doi: 10.1111/j.1365-2966.2012.20951.x

Blake, G. A., Sutton, E. C., Masson, C. R., & Phillips, T. G. 1987, The Astrophysical Journal, 315, 621, doi: 10.1086/165165

Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, The Astrophysical Journal, 547, 792, doi: 10.1086/318388

Duarte-Cabral, A., Colombo, D., Urquhart, J. S., et al. 2021, The SEDIGISM survey: molecular clouds in the inner Galaxy, Monthly Notices of the Royal Astronomical Society, Volume 500, Issue 3, pp.3027-3049, doi: 10.1093/mnras/staa2480

Harvard CfA, S. 2023, 1.2m CO Survey Archive, https://lweb.cfa.harvard.edu/rtdc/CO/

Heyer, M., & Dame, T. M. 2015, Annual Review of Astronomy and Astrophysics, 53, 583, doi: 10.1146/annurev-astro-082214-122324

Mangum, J. 2006. https://safe.nrao.edu/wiki/pub/Main/RadioTutorial/radio-obs-modes.pdf

Mangum, J. G., Emerson, D. T., & Greisen, E. W. 2007, Astronomy & Astrophysics, 474, 679, doi: 10.1051/0004-6361:20077811

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, Astronomy & Astrophysics, 571, A13, doi: 10.1051/0004-6361/201321553

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, Astronomy & Astrophysics, 641, A1, doi: 10.1051/0004-6361/201833880

Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020b, Astronomy & Astrophysics, 641, A4, doi: 10.1051/0004-6361/201833881

Schuller, F., Urquhart, J. S., Csengeri, T., et al. 2021, Monthly Notices of the Royal Astronomical Society, 500, 3064, doi: 10.1093/mnras/staa2369

Stanimirović, S. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 278, Single-Dish Radio Astronomy: Techniques and Applications, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, 375–396, doi: 10.48550/arXiv.astro-ph/0205329

Walker, C. K. 2015, Terahertz Astronomy (CRC Press)

A Coordinates and emission of the first set of candidates

Table 1: Emission of the potential OFF-positions observed by ALMA, selected with the first method from the SEDIGISM data-cubes. The only clean candidate appears to be OFFJ1328-6209.

Name	RMS (mk)	Peak (mk)	SN	RA (deg)	DEC (deg)
OFFJ1744-2811	19.19	190.12	9.91	266.1506415	-28.1885255
OFFJ1743-2912	40.13	486.27	12.12	265.9832040	-29.2039634
OFFJ1745-2830	36.84	200.12	5.43	266.2562445	-28.5116424
OFFJ1744-2934	145.06	684.06	4.72	266.1631080	-29.5789906
OFFJ1744-2935	97.67	517.77	5.30	266.1915645	-29.5941635
OFFJ1745-2933	75.63	686.64	9.08	266.2954470	-29.5628444
OFFJ1745-2931	158.05	941.33	5.96	266.3088360	-29.5235687
OFFJ1745-2932	103.34	765.73	7.41	266.3112750	-29.5403325
OFFJ1746-2858	97.73	543.36	5.56	266.5639980	-28.9762291
OFFJ1747-2815	108.52	510.20	4.70	266.7952425	-28.2504805
OFFJ1747-2812	139.09	914.48	6.57	266.8082775	-28.2111563
OFFJ1746-2847	201.93	1610.70	7.98	266.6386830	-28.7874447
OFFJ1746-2919	60.65	303.79	5.01	266.5136685	-29.3325736
OFFJ1747-2815	151.24	729.33	4.82	266.8090635	-28.2609423
OFFJ1746-2934	32.37	198.99	6.15	266.5165065	-29.5720076
OFFJ1748-2812	127.04	768.99	6.05	267.1426125	-28.2133600
OFFJ1748-2920	8.34	97.21	11.65	267.1007745	-29.3433347
OFFJ1750-2846	8.67	131.21	15.14	267.7200015	-28.7785255
OFFJ1332-6207	212.44	3619.11	17.04	203.2210425	-62.1213849
OFFJ1328-6209	6.13	20.64	3.37	202.2203820	-62.1534597
OFFJ1334-6227	22.12	423.95	19.17	203.5640205	-62.4561484
OFFJ1328-6228	11.50	178.49	15.52	202.0240635	-62.4812958
OFFJ1318-6229	10.12	97.28	9.62	199.7201655	-62.4999781
OFFJ1326-6237	72.89	1193.57	16.38	201.6929925	-62.6279948
OFFJ1332-6242	81.17	1388.14	17.10	203.0725560	-62.7071207
OFFJ1330-6245	8.63	68.62	7.95	202.7352780	-62.7607837
OFFJ1324-6258	20.90	324.88	15.54	201.1868535	-62.9678405
OFFJ1322-6241	8.67	69.88	8.06	200.7365970	-62.6879474